

How predatory journals leak into PubMed

Andrea Manca PhD, David Moher PhD, Lucia Cugusi, PhD, Zeevi Dvir PhD, Franca Deriu MD PhD

■ Cite as: *CMAJ* 2018 September 4;190:E1042-5. doi: 10.1503/cmaj.180154

Recent reports that PubMed, one of the world's leading biomedical databases, includes predatory journals and their publications^{1,2} is cause for concern. PubMed handles millions of queries daily and represents a key source of knowledge for health researchers worldwide. Much medical research that underpins clinical practice relies on the findings generated by peer-reviewed studies that are retrieved via biomedical databases, in particular, those that are free to search such as MEDLINE and PubMed. Thus, it is imperative that these databases are free of contamination by the outputs of predatory journals with their critically flawed peer review procedures.³ We analyze why this is happening and identify some possible solutions to stop the penetration of predatory journals and publishers into biomedical databases.

What is a predatory journal and why does predatory publishing matter?

There is no clear consensus definition for predatory publishers and journals. Such journals have been referred to as “low quality, amateurish, and often unethical academic publishing that is usually Open Access (OA).”⁴ Although the descriptor “predatory” has been criticized,⁵ it is now widely recognized and accepted. The term was introduced by librarian Jeffrey Beall⁶ who also created and maintained updated blacklists of potential predatory publishers and journals, which were acclaimed by many for their utility in helping scholars to spot fraudulent journals until they were removed in January 2017 for undisclosed reasons.⁷ Recently, there has been a move to favour identification based on potentially predatory publishing practices.⁸ In this regard, a set of evidence-based salient features of journals suspected to be predatory has been proposed; these features are straightforward to assess (Box 1).⁹ A recent commentary identified potential characteristics of predatory journals related to categories such as journal operations, editorial and peer review, manner of communication with authors, article processing charges, and ways of dissemination, indexing and archiving.¹⁰ Along with the development of general criteria to discern legitimate from predatory operations, the scientific community has shown an increasing demand for whitelists of journals, because it seems more feasible to evaluate journals objectively based on what they do rather than what they do not do. The Directory of Open Access Journals is widely considered the most comprehensive whitelist of open-

KEY POINTS

- PubMed, MEDLINE and PubMed Central are all funded by the National Library of Medicine but are different databases.
- PubMed has been reported to include some articles published in predatory journals.
- MEDLINE and PubMed policies for the selection of journals for database inclusion are slightly different.
- Weaknesses in the criteria and procedures for indexing journals in PubMed Central may allow publications from predatory journals to leak into PubMed.
- Closing these loopholes is necessary to protect the integrity of reputable databases and safeguard biomedical research.

access journals deemed to be reputable.¹¹ Yet, adequate quality control of journal whitelists remains a problem.¹²

Most predatory journals are active in the biomedical sphere.¹³ Predatory publishing practices allow bad research that is poorly peer-reviewed, or published without peer review, to be published alongside real science that is rigorously reviewed, thus obscuring scientific truth.¹⁴ Furthermore, predatory publications may be included in the resumés of scholars seeking employment or promotion and tenure.¹⁵

What are MEDLINE and PubMed policies for the inclusion of journals?

Both MEDLINE and PubMed are maintained by the National Library of Medicine, the world's largest biomedical library, founded in 1836 within the National Institutes of Health.¹⁶ Among the library's products, three — MEDLINE, PubMed and PubMed Central — play a prominent role in the access to and distribution of scientific knowledge.

These three databases overlap considerably, as shown in Figure 1. Users of the PubMed search engine query all three databases simultaneously as the default option. However, each database can be searched separately, because there are some important differences between them.

MEDLINE is the largest subset of PubMed; it started in print in the 1960s and contains more than 24 million references to journal articles in life sciences with a concentration on biomedicine, the largest fraction of which are included as PubMed references.¹⁷

A distinctive feature of a MEDLINE literature search is that the records are indexed with National Library of Medicine Medical Subject Headings (MeSH), the vocabulary thesaurus controlled by the National Library of Medicine that is used for indexing articles for PubMed.

PubMed was introduced in 1996 as an experimental database under the Entrez retrieval system; free full access to MEDLINE

Box 1: Salient characteristics of potential predatory journals

- The scope of interest includes nonbiomedical subjects with biomedical topics.
- The website contains errors in spelling and grammar.
- Images are distorted or fuzzy, intended to look like something they are not or are unauthorized.
- The language on the home page targets authors.
- The Index Copernicus Value is promoted on the website.
- There is no description of the process for handling manuscripts.
- The website requests that manuscripts should be submitted via email.
- Rapid publication is promised.
- There is no retraction policy.
- Information on whether and how journal content will be digitally preserved is absent.
- The processing or publication charge for the article is low (e.g., < US\$150.00).
- Journals that claim to be open access either retain copyright of the published research or fail to mention copyright.
- The contact email address is nonprofessional and not affiliated with a journal (e.g., @gmail.com or @yahoo.com).

Note: Adapted from Shamseer and colleagues.⁹

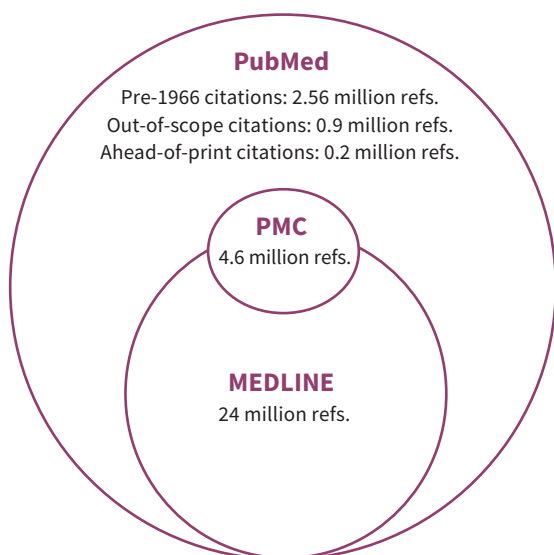


Figure 1: Venn diagram showing the overlap and unique features of PubMed and MEDLINE. References provided as of Sept. 30, 2017. Note: PMC = PubMed Central, refs. = references.

started from 1997. PubMed contains more than 28 million citations and abstracts from journals indexed in MEDLINE and, unlike MEDLINE, also life science journals and online books, which are free to access through the National Center for Biotechnology Information Bookshelf. PubMed also provides access to journals and manuscripts deposited in PubMed Central, additional relevant websites and links to the other molecular biology resources at the National Center for Biotechnology Information.

PubMed Central is an electronic archive of full-text journal articles that offers free access to its content. It contains more than 4 million articles, most of which have a corresponding entry in PubMed, and a small portion of which do not because they fall outside of PubMed's scope (e.g., book reviews).¹⁸

MEDLINE and PubMed have different policies for journal selection, which means that just over 5600 journals are indexed in the former and about 30 000 are indexed in the latter (as of March 2018). The decision on whether to index a journal in MEDLINE is made by the Director of the National Library of Medicine based on quality and policy considerations that are set by the Board of Regents of the library. A Literature Selection Technical Review Committee is then responsible for reviewing journal titles and assessing the quality of their content. This committee comprehensively reviews applications through a multistep procedure.¹⁹ To cross-check a journal's data, the procedure follows guidelines from the International Committee of Medical Journal Editors that were developed collaboratively by the World Association of Medical Editors, the Committee on Publication Ethics, the Directory of Open Access Journals and the Open Access Scholarly Publishers Association, and are based on principles of transparency that represent a minimum set of criteria to meet and maintain when seeking membership in any of the four organizations.²⁰

The strict processes and criteria outlined above for a journal's inclusion in MEDLINE are not mandatory for inclusion in PubMed and PubMed Central. Although open access journals and publishers indexed in MEDLINE are retrievable in the Directory of Open Access Journals and the Open Access Scholarly Publishers Association, a worrying number of those indexed by PubMed are not listed in either directory, because PubMed policies for journal selection are less strict. As a result, PubMed may well include journals that either did not apply for MEDLINE indexing, or did apply but did not satisfy minimum criteria for inclusion.

Other factors, such as the need to deposit research funded by the National Institutes of Health under the Public Access Policy, the broader scope and the wide range of available material in PubMed compared with MEDLINE, may contribute to the absence of many PubMed citations in the whitelists of the Directory of Open Access Journals and the Open Access Scholarly Publishers Association.

What are the potential loopholes?

The journal selection policy for PubMed Central comprises a pre-application — a requirement for an ISSN number, which is easy to obtain;²¹ a two-year history of publication; and a minimum of 25 peer-reviewed articles — followed by a six-step procedure

(i.e., Submit Application, Initial Application Screening, Scientific Quality Review, Technical Evaluation, Pre-Production and Release to Live).²² This is a looser and apparently more flexible procedure than MEDLINE's; it is supervised by the National Library of Medicine, which also provides a qualitative review. Although the preapplication phase does consider conformance with guidelines and best practices published by professional organizations, including Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals from the International Committee of Medical Journal Editors,²³ Principles of Transparency and Best Practice in Scholarly Publishing²⁰ and the Recommended Practices for the Presentation and Identification of E-Journals from the National Information Standards Organization,²⁴ this process is not always uniform. Some journals indexed in PubMed Central are less than two years old and with fewer than 25 articles (often fewer than 10, even after counting the total number of articles issued on the journal's website), which can be easily verified by searching the journal's or publisher's name in the National Library of Medicine's catalog (www.ncbi.nlm.nih.gov/nlmcatalog/) to check the publication start year and the number of articles published in that journal.

The National Library of Medicine may consider an application from a publishing organization that has been publishing scholarly content for less than two years if there is evidence that the management and individuals responsible for editorial quality and operations have adequate experience in comparable positions at other organizations. That is, any journal or publisher candidate that did not apply for or meet criteria can be indexed in PubMed Central if they show that at least one credible published author is listed among its editorial board. However, it has been reported that predatory journals often use the identities of reputable researchers, who are unaware of being listed in the editorial board of these journals.²⁵ This may explain the unauthorized use of academic names for editorial boards and the relentless number of daily invitations many researchers receive to join an editorial board of a never-heard-of journal.

In its qualitative review of candidate journals to be considered for inclusion in PubMed Central, the National Library of Medicine considers the 16 recommendations outlined in the guideline from the International Committee of Medical Journal Editors²³; however, in the absence of any assessment from MEDLINE's Literature Selection Technical Review Committee, the journal's title and publisher's name are not double-checked with the lists of these directories. Predatory journals often register journal names very similar to established legitimate journals and may not be recognized as fraudulent. As a result, open-access journals that do not qualify for inclusion may be indexed in PubMed Central and subsequently "leak" into the PubMed database.

How could predatory journals be kept out of PubMed?

The National Library of Medicine could raise the bar for journal inclusion in PubMed and PubMed Central by requiring all candidates to satisfy all three MEDLINE preapplication requirements

fully (application of quality and policy considerations set by the Board of Regents of the NLM; review of journals' titles and quality by the Literature Selection Technical Review Committee; and cross-check of journals' data using minimum criteria set out in 2014 International Committee of Medical Journal Editors guidelines) and by adding a fourth criterion of asking candidate journals and publishers to show membership in the Directory of Open Access Journals, Open Access Scholarly Publishers Association, Committee on Publication Ethics and World Association of Medical Editors. Double-checking of journal titles and publishers by the organization's Literature Selection Technical Review Committee, which is mandatory for inclusion in MEDLINE, would help identify those journals that slip into PubMed via PubMed Central despite not being whitelist journals. Additionally, the National Library of Medicine could make information about journal applicants available for scrutiny by the wider research community. Collectively, these safeguards might constrain the infiltration of PubMed by predators and the use of the National Library of Medicine in the inadvertent promotion of predatory journals.

References

- Manca A, Martinez G, Cugusi L, et al. Predatory open access in rehabilitation. *Arch Phys Med Rehabil* 2017;98:1051-6.
- Manca A, Martinez G, Cugusi L, et al. The surge of predatory open-access in neurosciences and neurology. *Neuroscience* 2017;353:166-73.
- Manca A, Cugusi L, Dvir Z, et al. PubMed should raise the bar for journal inclusion. *Lancet* 2017;390:734-5.
- Berger M. *Everything you ever wanted to know about predatory publishing but were afraid to ask*. Chicago: American Library Association; 2017. Available: www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2017/EverythingYouEverWantedtoKnowAboutPredatoryPublishing.pdf (accessed 2018 June 24).
- Eriksson S, Helgesson G. Time to stop talking about 'predatory journals'. *Learn Publ* 2018;31:181-3.
- Beall J. Predatory publishers are corrupting open access. *Nature* 2012;489:179.
- Strielkowski W. Predatory journals: Beall's List is missed. *Nature* 2017;544:416.
- Butler D. Investigating journals: the dark side of publishing. *Nature* 2013;495:433-5.
- Shamseer L, Moher D, Madueke O, et al. Potential predatory and legitimate biomedical journals: can you tell the difference? A cross-sectional comparison. *BMC Med* 2017;15:28.
- Moher D, Shamseer L, Cobey KD, et al. Stop this waste of people, animals and money. *Nature* 2017;549:23-5.
- Van Noorden R. Open-access website gets tough. *Nature* 2014;512:17.
- Bohannon J. Who's afraid of peer review? *Science* 2013;342:60-5.
- Moher D, Srivastava A. You are invited to submit... *BMC Med* 2015;13:180.
- Beall J. Medical publishing triage — chronicling predatory open access publishers. *Ann Med Surg (Lond)* 2013;2:47-9.
- Bagues MF, Sylos-Labini M, Zinovyeva N. *A walk on the wild side: an investigation into the quantity and quality of "predatory" publications in Italian academia*. LEM Working Paper Series, No. 2017/01. Pisa (Italy): Scuola Superiore Sant'Anna, Laboratory of Economics and Management (LEM); 2017.
- About the National Library of Medicine. Bethesda (MD): U.S. National Library of Medicine; 2004 Apr. 22 (updated 2018 June 22). Available: www.nlm.nih.gov/about/index.html (accessed 2018 June 24).
- MEDLINE, PubMed, and PMC (PubMed Central): How are they different? Bethesda (MD): U.S. National Library of Medicine; 2018 Apr. 20 (updated 2018 June 22). Available: www.nlm.nih.gov/bsd/difference.html (accessed 2018 June 24).
- What is the difference between PubMed Central (PMC) and PubMed? Bethesda (MD): U.S. National Library of Medicine. Available: <https://support.nlm.nih.gov/knowledgebase/article/KA-03247/en-us> (accessed 2018 June 24).

19. MEDLINE[®] journal selection [fact sheet]. Bethesda (MD): U.S. National Library of Medicine; 1988 Jan. 1 (updated 2018 Mar. 14). Available: www.nlm.nih.gov/pubs/factsheets/jsel.html (accessed 2018 June 24).
20. Principles of transparency and best practice in scholarly publishing. Hampshire (UK): Committee on Publication Ethics (COPE); 2018. Available: <https://publicationethics.org/resources/guidelines-new/principles-transparency-and-best-practice-scholarly-publishing> (accessed 2018 June 24).
21. Requesting an ISSN. Paris: International Standard Serial Number (ISSN) International Centre. Available: www.issn.org/services/requesting-an-issn/ (accessed 2018 June 24).
22. How to participate in PMC. Bethesda (MD): U.S. National Library of Medicine; (updated 2018 June 12). Available: www.ncbi.nlm.nih.gov/pmc/pub/addjournal/ (accessed 2018 June 24).
23. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. International Committee of Medical Journal Editors (ICMJE); (updated December 2017). Available: www.icmje.org/icmje-recommendations.pdf (accessed 2018 June 24).
24. PIE-J: the presentation & identification of e-journals — A recommended practice of the National Information Standards Organization. Baltimore: National Information Standards Organization (NISO); 2013. Available: https://groups.niso.org/apps/group_public/download.php/10368/rp-16-2013_pie-j.pdf (accessed 2018 June 24).
25. Sorokowski P, Kulczycki E, Sorokowska A, et al. Predatory journals recruit fake editor. *Nature* 2017;543:481-3.

Competing interests: None declared.

This article has been peer reviewed.

Affiliations: Department of Biomedical Sciences (Manca, Deriu), University of Sassari, Sassari, Italy; Centre for Journalology (Moher), Ottawa Hospital Research Institute, Ottawa, Ont.; Department of Medical Sciences and Public Health (Cugusi), University of Cagliari, Cagliari, Italy; Department of Physical Therapy (Dvir), Sackler Faculty of Medicine, University of Tel Aviv, Tel Aviv, Israel

Contributors: Andrea Manca, Lucia Cugusi and Franca Deriu drafted the manuscript.

Andrea Manca conceived and designed the study, gathered data, and contributed to the analysis and interpretation of the data. David Moher contributed to the conception of the study, interpreted data and revised the manuscript critically for important intellectual content. Lucia Cugusi co-conceived the study, and contributed to the analysis and interpretation of the data. Zeevi Dvir contributed to the design of the study, interpreted data and revised the manuscript critically for important intellectual content. Franca Deriu conceived and designed the study, gathered data, and contributed to the analysis and interpretation

of the data. All of the authors gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding: This work did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. David Moher is funded by a University Research Chair at the University of Ottawa, Ottawa, Ont.

Correspondence to: Franca Deriu, deriuf@uniss.it